自然语言处理技术赋能新闻生产自动化应用案例探研

张建伟
（新华社北美总分社）

摘 要：时效性是衡量新闻的决定性标尺，它要求新闻媒体在新闻发生时要用最快的速度对新闻事件进行真实、准确地报道。为了能够在日趋激烈的新闻竞争中处于不败之地，新闻机构广泛利用人工智能技术，特别是自然语言处理技术，力图实现新闻制作自动化。本文主要介绍新闻媒体利用自然语言处理技术实现新闻自动化的两个案例，并探讨新华社北美总分社在自然语言处理技术当前应用的情况，对未来发展提出了的一些建议。

关键词：自然语言处理；新闻媒体；聊天机器人；自动化；人工智能　　**中图分类号：**G210.7　　**文献标识码：** A
文章编号：1671-0134（2022）04-142-03　　DOI：10.19483/j.cnki.11-4653/n.2022.04.042
本文著录格式：张建伟 . 自然语言处理技术赋能新闻生产自动化应用案例探研 [J]. 中国传媒科技，2022（04）：142-143，157.

导语

随着大数据、人工智能技术的发展，新闻行业迎来了历史性的变革。新闻机构利用人工智能和大数据技术不断推动新闻业态的发展。从写稿机器人到个性化新闻推荐，新技术给新闻业注入了新的能量。

特别是在自然语言处理（NLP）技术领域，新闻媒体在这方面进行了较多的探索。比如，NLP 技术对众包的数据进行清洗和预处理。或在新闻编辑环节使用自然言语生成（NLG）技术优化新闻制作；使用推荐算法个性化新闻推送；或使用聊天机器人（chat bot）与用户的实时互动。

1. 自动化图表新闻

根据相关新闻机构的调查发现，年轻读者一方面对涉及公共服务且具有重要价值的新闻故事表现出浓厚的兴趣；另一方面，对采用大段文字叙事的新闻则缺少耐心逐段阅读。为此，国外新闻机构希望通过采用漫画 + 文字的形式来发布新闻信息，通过此形式吸引更多的读者。[1]

虽然图表、漫画新闻表现形式并非首创，在新华社摄影新闻报道中已经是普遍采用的一种新闻形式。但图表新闻的制作过程对编辑仍存在不小的技术挑战：比如，如何快速、准确的挑选适合新闻主题的素材；如何将各种矢量图片组合并制作成最终的漫画模板。上述这些工作，即使对经验丰富的编辑也是一项费时的工作。能否将图表新闻通过计算机自动完成，提高新闻制作效率，减轻编辑的工作压力，就成为自动化图表新闻主要解决的问题。

有新闻媒体曾启动了为期 6 个月的一项实验，实验的主要目的就是尝试通过计算机自动制作图表新闻。第一期实验为期 6 周，主要是验证技术可行性。据介绍，为了降低项目复杂度，项目组将问题聚焦在公共健康方面的新闻，主要因为该类新闻大多具有相似的结构，多数内容是关于医学方面的新发现或新疗法。他们首先了解公共健康新闻的结构、重要的新闻构成要素。在此基础上，项目组对新闻制作流程进行了分解。[2]

项目组首先利用自然语言处理技术中相对成熟的“命名实体识别（NER）”技术来鉴别文字稿件中的关键实体信息；再将算法识别出来的实体信息作为参数，调取图片检索技术查找适合的漫画素材，通过预制的模板将文字新闻和漫画背景渲染成一幅漫画。这个原型服务通过网页应用的形式开放给编辑使用。编辑通过该网页应用，可以将精力集中于文字编辑，系统后台完成漫画搜索和模板渲染的工作，并将最终结果呈现给编辑，供编辑做进一步的调整和完善。

根据项目一期的效果 ，项目组启动了第二阶段的实验，仍然是 6 周时间。这一阶段，项目组着重解决漫画自动化生成的准确度和效率问题。为此项目组设计了一套自然语言处理的流水线，并对 NER 信息提取的算法进行了改进。

由于该系统主要利用 NER 技术来解析新闻稿件中的关键信息，此环节的准确程度决定了后续漫画模板的选择是否准确合理。为了使第一步关键信息提取更加准确，项目组尝试了多种方式，包括提取一组相关的实体信息名，让记者选择最相关的关键字并保留下来作为图片检索的参数。还尝试直接允许编辑输入关键字。

为了试图让系统自动挑选出最适合的图片作为漫画新闻素材，项目组根据医疗健康新闻特点，设计了一套规则。比如，如果在文字稿件中检测到数字信息，系统会生成一个简单的数据可视化图；如果编辑输入了引言，系统会将引言自动添加到图片对话框，并标注被引用者的信息。

然而，在这一阶段的实验中，项目组发现理想与现实的差距，比如：医学领域经常会发布一些有关疑难疾病的创新疗法，此时利用自动化的系统很难将 NER 信息给出对应的漫画素材相映射，找到合适素材或模板。为此，项目组邀请认知科学家、漫画理论家共同参与设计新闻

模板，利用更加通用的漫画模板尽量完善系统自动化流程。

2020 年 10 月，项目组使用该系统制作了一条有关新冠疫情的新闻，并发布在脸书和 Instagram 的“限时动态”中。目前该原型产品已经应用在媒体内部的健康新闻小组。利用该系统，编辑仅专注于完成文字编写工作，系统完成新闻主题的识别，根据识别出的主题搜索恰当的漫画模板，并进行合成渲染。

2. 利用聊天机器人发布新冠疫情信息

新冠疫情在很大程度上改变了人们生活、工作的方式。特别是在 2020 年疫情期间，对病毒的恐惧，迫使人们不得不选择大部分时间居家办公。与此同时，铺天盖地的虚假新闻和错误的信息，也使人们变得无所适从。

对于媒体来说，让受众在第一时间获取到真实、可靠的信息，自然成为新闻机构追求的首要目标。特别是在疫情肆虐的时期，人们对各种资源的需求更加迫切。为用户提供一条畅通且可靠的疫情信息播报通道，也就成为各新闻机构要解决的一个重要问题。

另外，随着社交媒体平台的不断发展，像微信、微博已经成为新闻消费者的重要社交网络平台。据路透社的一个统计报告，在 2018 年至 2020 年期间，大约有 1/3 的国外用户通过脸书获取新闻，部分国家甚至更高。社交媒体成为人们获取新闻的主要途径。[3]

正因为社交媒体在新闻传播方面的巨大影响力，有媒体选择尝试将聊天机器人与脸书的聊天应用（Facebook Messenger）相结合，发布新闻信息。一方面，疫情期间，记者可以在家方便地更新相关信息；另一方面，随着人工智能技术的进步，聊天机器人解决方案也已成熟；而且，根据脸书的统计，脸书的聊天客户端在疫情开始阶段的使用频率呈现爆发式增长。基于上述原因，将聊天机器人（chat bot）集成到脸书的聊天应用程序，或许是个不错的实践。[3]

该新闻机构尝试使用一款名为“德克斯特”（Dexter）的第三方工具作为聊天机器人的引擎。借助脸书的 Messenger 应用通道，记者除了发布最新的疫情感染数字以及相关的封城措施外，同时推送其他附加内容，以防止人们错过重要的新闻资讯。机器人可与用户进行实时互动，根据用户输入的内容，为用户提供相关的视频新闻或有关医护工作人员那些振奋人心的感人事迹。[4]

该新闻机构首先在东南亚部署了测试版程序，用于获取用户和机器人的交互情况。三周后，项目组增加了订阅功能，允许系统自动推送信息。根据项目组收集到的数据显示：大约 1/3 曾与机器人交互过的用户选择订阅该服务。该项目在越南和孟加拉两个国家进行了部署，这两地的新闻团队负责每天更新最新的疫情数据。这两组机器人总计收发大约一百万条交互信息。这一交互数量远高于项目组的预期。

根据用户与聊天机器人的交互情况，项目组发现：当新闻发布机构自动推送的消息对用户来说是重要话题时，用户的订阅意愿十分强烈。另外，脸书的收件箱是一个未被充分利用的空间，特别是对于聊天机器人等智能化交互，这一通道是一个成熟交互途径，也是可以充分开发的交互方式。但同时，如何让用户发现聊天机器人这一功能，确实是一个挑战。目前项目组使用的方法就是通过广告宣传，通过在各种渠道宣传，让用户知道并发现这个服务。另外，在 2020 年的秋天，该新闻机构尝试在美国大选期间利用机器人推送相关消息，并与用户互动，提供详细的选情报道，采用的也是类似的解决方案。

3. 对新闻行业技术创新的启示

上述两个项目是国外新闻机构将人工智能技术与新闻生产和传播相结合的典型案例，这其中既有利用计算机辅助编辑、记者提高新闻播发效率；也有通过新技术增强与用户互动，获取用户行为数据，进一步指导新闻编发工作。

笔者所在单位，在人工智能应用方面进行了一些探索，诸如在编辑部使用的新一代采编系统中囊括了诸多人工智能工具，比如：智能检校、语音识别、智能辅助翻译、图片文字识别等，都是利用自然语言处理技术实现工作的自动化。这些实践在实际工作中已经发挥了重要的作用，有效提高了新闻制作效率，节约了时间，获得编辑、记者的好评。有记者反馈，利用语音识别工具，可极大提高采访素材处理效率，原来需要几个小时才能处理完成的文字素材，可通过 AI 工具在几分钟内即可完成初加工，记者可以将更多的精力投入到深度内容的挖掘中。

由此可见，人工智能技术能够有效提升新闻制作效率，因此应该充分利用这一技术优势，将人工智能技术与新闻制作、传播更多的进行融合。根据不同的应用场景或新闻形式，设计相应的新闻制作工作流引擎，将原始素材的初加工、查询工作自动化，让编辑将更多的精力放在新闻故事的创作上。让新闻采编工作全流程进入智能化、自动化的快车道。

其次，随着技术不断进步，新闻行业也正在经历翻天覆地的变革，各种创新层出不穷。新华社作为国家通讯社，如何履行好党中央“喉舌”“耳目”职能，发挥正确引导国内舆论、积极影响国际舆论的重要作用，是新闻工作者同样也是技术创新的初心和使命。与此同时，也不能忽视的是在当下这个时代，新闻创新可能不再是开创单一的传播渠道，而是以为多样化的受众提供真正有价值的数字新闻体验为目标。数字媒体生态系统的加速变化要求新闻从业者改变传统的制作、分发和传播新闻的方式。如何在完成国家通讯社的职能和使命的同时，为多样的用户提供个性化的新闻产品，也是我社面临的

（下转第157页）